Gene, 161 (1995) 183–187

Attach #22
Exhibit C

G NE 08941

Increased exon-trapping efficiency through modifications to the pSPL3 splicing vector

(Gene identification; cloning; exon amplification; human)

Timothy C. Burn, Timothy D. Connors, Katherine W. Klinger and Gregory M. Landes

Department of Human Genetics, Integrated Genetics Inc., Framingham, MA 01701-9332, USA

Received by J.W. Messing: 29 November 1994; Revised/Accepted: 21 February 1995; Received at publishers: 20 March 1995

SUMMARY

Exon trapping allows for the rapid identification and cloning of coding regions from cloned eukaryotic DNA. In preliminary experiments, we observed two phenomena which limited the exon-trapping efficiency of pSPL3-based systems. The first factor that affected performance was revealed when we found that up to 50% of the putative trapped exons contained sequences derived from the intron of the pSPL3 trapping vector. Removal of the DNA sequences responsible for the cryptic splice event from the original splicing vector resulted in a new vector, pSPL3B. We demonstrate that pSPL3B virtually eliminates pSPL3-only spliced products while maximizing the proportion of exon traps containing genomic DNA (>98%). The other step which impacted performance was our observation that a majority of the ampicillin-resistant (Ap^R) clones produced after shotgun subcloning from Ap^R cosmids into pSPL3 were untrappable, pSPL3-deficient, recircularized cosmid vector fragments. Replacement of the pSPL3 Ap^R gene with the Cm^R cassette encoding chloramphenicol (Cm) acetyltransferase enabled selection for only pSPL3-containing Cm^R clones. We show a 30–40-fold increase in the initial subcloning efficiency of cosmid-derived fragments with pSPL3-CAM, when compared to pSPL3. The collective vector alterations described improve the overall exon-trapping efficiency of the pSPL3-based trapping system.

INTRODUCTION

The method of exon trapping has been recently described for the rapid and efficient isolation of coding regions from genomic DNA (Auch and Reth, 1990; Duyk et al., 1990; Buckler et al., 1991: Church et al., 1994). One of the more widely used exon-trapping procedure utilizes the pSPL3 plasmid which contains rabbit β-globin coding sequences separated by a portion of the HIV-*tat* gene (Buckler et al., 1991: Church et al., 1993; 1994). Genomic DNA (YAC, cosmid or P1 clones) is cloned into the intron of the *tat* gene and the resulting subclones are transfected into COS-7 cells. SV40 sequences in the vector allow for both replication of the pSPL3 containing DNAs and transcription of the cloned genomic DNAs. Exons within the subcloned genomic DNAs that are spliced into the globin-*tat* transcript are recovered by RT-PCR. A major advantage of exon trapping is that the expression of the subcloned DNA is directed by a viral promoter, thus developmental and tissue specific expression of genes is not a concern.

We have initiated studies to assemble a transcriptional

Correspondence to: Dr. T.C. Burn, Department of Human Genetics, Integrated Genetics, Inc., One Mountain Road, Framingham, MA 01701-9332, USA. Tel. (1-508) 872-8400, ext. 2224; Fax (1-508) 620-1203; e-mail: gmilleriglab@delphi.com

Abbreviations: Ap, ampicillin; bp, base pair(s); CAT, Cm acetyltransferase; Cm, chloramphenicol; EtdBr, ethidium bromide; HIV, human immunodeficiency virus; kb, kilobase(s) or 1000 bp; LB, Luria-Bertani medium; MCS, multiple cloning site(s); nt, nucleotide(s); oligo, oligodeoxyribonucleotide; *ori*, origin(s) of replication; PCR, polymerase chain reaction; Pollk, Klenow (large) fragment of *E. coli* DNA polymerase I; R, resistant/resistance; RT, reverse transcriptase; SDS, sodium dodecyl sulfate; SSC, 0.15 M NaCl/0.015 M Na_3-citrate pH 7.6.

SSDI 0378-1119(95)00223-5

map for a 750-kb CpG-rich region of human chromosome 16p13.3 using exon trapping. This region of chromos me 16 is of particular interest since it contains the gene responsible for greater than 90% of all cases of autosomal dominant polycystic disease (APKD; Reeders et al., 1985; Germino et al., 1992; Somlo et al., 1992; The European Polycystic Kidney Disease Consortium, 1994). It also contains the disease gene for tuberous sclerosis (TSC2; The European Chromosome 16 Tuberous Sclerosis Consortium, 1993) and possibly the gene responsible for familial Mediterranean fever (MEF; Aksentijevich et al., 1993). We have proceeded to use exon trapping to identify coding sequences from the PKD-1 locus. In our initial experiments, we observed that a significant percentage of the trapped products contained a pSPL3-derived cryptic exon from within the HIV-*tat* intron. We also observed a poor subcloning efficiency in the initial shotgun subcloning of genomic cosmid clones into the pSPL3 vector, with the majority of the ligation products being recircularized cosmid vector. We report the identification of a cryptic exon within the pSPL3 vector and demonstrate the utility and efficiency of a modified pSPL3 vector lacking the cryptic splice sites. We also demonstrate the utility of a modified pSPL3 vector in which the Ap^R gene has been replaced by the Cm^R gene.

EXPERIMENTAL AND DISCUSSION

(a) The pSPL3 splicing vector contains a cryptic exon

Exon trapping, as described by Buckler et al. (1991) and Church et al. (1993; 1994), identifies secondary PCR products larger than the 177-bp product generated by the pSPL3 vector alone. In our initial studies, we found that up to 50% of our putative trapped exons were of a single size (approx. 120-bp longer in size than the 177-bp pSPL3 product) and were not derived from the parental P1 or cosmid inserts. The prevalence of the unexplained exon was independent of the initial subcloning efficiency and solely dependent on pSPL3, which alone yielded the cryptic exon in up to 50% of the cases (Fig. 1, lane 2). Characterization of the cryptic splice product indicated that it contained 116 nt of sequence from the intronic portion of pSPL3 (Fig. 2). Sequences 5′ of the cryptic exon were found to contain a pyrimidine-rich tract and a 5′ splice acceptor site with some homology to the published consensus sequence (Mount, 1982), while the 3′ splice donor site of the cryptic exon contained sequences which are nearly identical to the published consensus sequence (Mount, 1982).

To increase the efficiency and reliability of exon trapping, we removed the cryptic exon using convenient

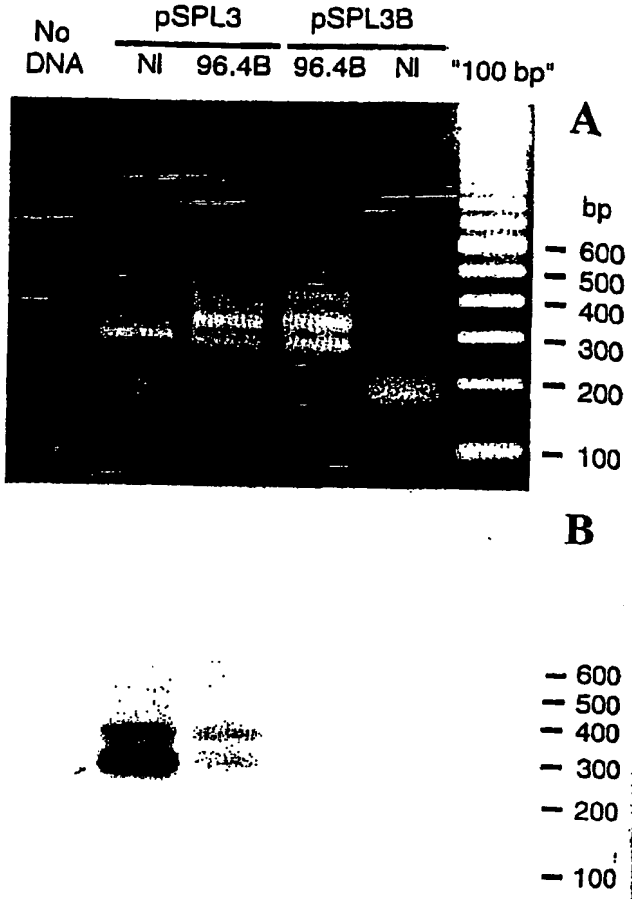

Fig. 1. Comparison of the pSPL3 and pSPL3B trapping vectors. (A) EtdBr-stained gel of PCR products from an exon-trapping experiment using vector without insert (NI) or vector into which genomic DNA from the 96.4B bacteriophage P1 clone had been shotgun subcloned (96.4B). A PCR reaction in which no template was added (No DNA) and a 100-bp size ladder are also shown. The 177-bp product in the pSPL3B NI (no insert) lane is the globin-*tat* RT-PCR product without an internal trapped exon. (B) The same gel was blotted to nylon membrane and hybridized with the HIV-*tat* intron. Methods: The 96.4B bacteriophage P1 clone was cut to completion with *Bam*HI + *Bgl*II and shotgun subcloned into dephosphorylated *Bam*HI-cut pSPL3 or pSPL3B. After overnight ligation at 16°C, the reactions were transformed into competent *E. coli* HB101. The following day, colonies were scraped off those plates having the highest transformation efficiency (based on comparison to 'no insert' control ligations). Alkaline lysis minipreps were performed on the plate scrape pools and the resulting miniprep DNA was electroporated into COS-7 as described (Church et al., 1994). At about 48 h post-electroporation, RT-PCR was performed as described by the manufacturer (Life Technologies, Gaithersburg, MD, USA). The resulting secondary PCR products were separated on a 1.5% agarose gel and blotted to nylon membrane. The HIV-*tat* intron was amplified using PCR with primers at nt 671–691 (5′-TGCACGCTCTAGAGTCG) and 3102–3116 (5′-ATACCCC-TCGGGAGATC) of pSPL3. The resulting PCR product was gel-purified and labeled by random priming (Feinberg and Vogelstein, 1983). The blot was hybridized at 65°C in 0.25 M Na-phosphate pH 7.2/7% SDS, and washed in 0.2 × SSC/0.1%SDS at 65°C for 30 min.

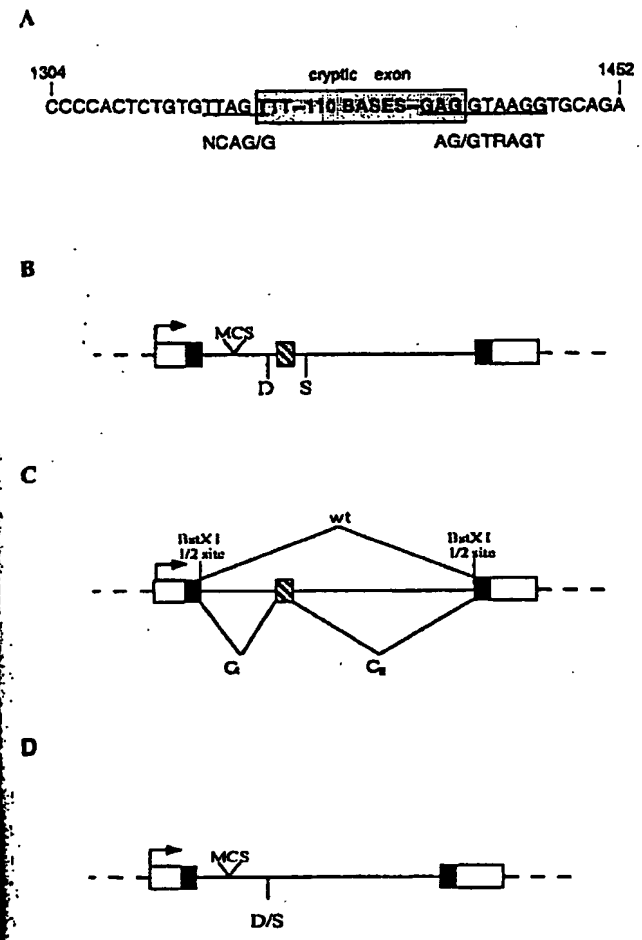

Fig. 2. Removal of the crytpic exon from the pSPL3 splicing vector. (A) The nt sequence for the cryptic exon's 5′ splice acceptor and 3′ splice donor sites. The numbers above the sequence indicate the position of the nt in the pSPL3 vector. The cryptic exon is indicated by the shaded box. Sequences with homology to the consensus splice sites (shown below pSPL3 sequence) are underlined. (B) Schematic diagram of the pSPL3 splicing vector. The SV40 promoter is indicated by an arrow. Open boxes indicate the globin coding sequences while the black boxes represent the HIV-*tat* splice donor and acceptor sequences. The solid black line designates the HIV-*tat* intronic sequences including the cryptic exon (hatched box). The position of the MCS, *Dra*III (D) and *Stu*I (S) sites are also shown. (C) The splicing reactions which lead to the wild-type (wt) and cryptic (C_1 and C_2) splice products are indicated. (D) Schematic diagram of the pSPL3B vector which lacks the cryptic exon. The cryptic exon from pSPL3 was removed by digesting pSPL3 (LifeTechnologies) with *Dra*III + *Stu*I. A 1-kb *Stu*I-*Dra*III fragment containing nt 321 to 1311 of pSPL3 was gel-purified, and the *Dra*III end was blunt-ended using PolIk. The resulting blunt-ended fragment was subcloned into pSPL3 which had been digested with *Stu*I to remove nt 321 to 1548. The resulting vector, pSPL3B, is identical to pSPL3 except for the absence of nt 1311 through 1548. Deletion of the cryptic exon was verified by nt sequence analysis.

restriction enzyme sites to generate pSPL3B (as described in the legend to Fig. 2). A side by side comparison of pSPL3 and pSPL3B was performed to demonstrate the utility of the modified vector. DNA from the genomic P1 clone 96.4B was shotgun subcloned into either pSPL3 or pSPL3B. The subcloning efficiency of the two reactions was equivalent based on transformation efficiencies and restriction digestion analysis (data not shown). Exon trapping was performed in parallel and the secondary PCR products were separated on an agarose gel and Southern blotted prior to subcloning. The removal of the cryptic exon did not decrease our ability to trap exons as judged by the number and intensity of bands on an EtdBr-stained agarose gel (Fig. 1A); however, hybridization with the HIV-*tat* intron from pSPL3 indicated that cryptic products were present in the pSPL3 reactions but not in the pSPL3B reactions (Fig. 1B). We also observed hybridization of the HIV-*tat* intron probe to a second less abundant product at 375 bp (Fig. 1B). The second product appears to result from the cryptic exon's use of a minor splice acceptor or donor site, since it is also eliminated by the 238-bp deletion in pSPL3B. These data indicate that the removal of the cryptic exon significantly reduces the levels of non-specific products in the secondary PCR reactions.

(b) Removal of the cryptic exon reduces the percentage of subcloned products containing HIV-*tat* sequences

To demonstrate that removal of the cryptic splice products decreases the percentage of non-specific products in the final cloning step, we hybridized the pSPL3 intron to colony lifts from a number of exon-trapping experiments. Exon trapping was performed on three P1 clones and colony hybridizations were performed on the resulting subcloned exons using the HIV-*tat* intron as a probe. The results from this comparison are summarized in Table I. With the original pSPL3 vector, 5–50% of all products obtained by exon trapping contained pSPL3 intronic sequences, while 0–2% contained cryptic products for pSPL3B. When only the experiments using P1 clones were considered, 22% of all the traps from the pSPL3 experiment were HIV-*tat* derived, while less than 1% were HIV-*tat* derived in the pSPL3B experiments. The products hybridizing to the HIV-*tat* intron in the pSPL3B experiments presumably represent a less efficiently spliced cryptic exon(s) which is still present in the vector

(c) Replacement of the pSPL3 Ap^R gene with the Cm^R gene significantly increases shotgun subcloning efficiency of cosmid clones

A pSPL3 derivative containing the Cm^R gene instead of the Ap^R gene has also been produced. As mentioned above, recircularized cosmid vector significantly reduced the overall efficiency of subcloning into the pSPL3 vector. Despite optimization of ligation conditions, as many as 95% of all Ap^R colonies were found to be cosmid derived

TABLE I

Comparison of pSPL3 and pSPL3B exon trapping vectors

Clone	Cryptic exons[a]	
	pSPL3	pSPL3B
No insert (Control)	203/400 (51%)	1/220 (0.4%)
96.4B	10/185 (5%)	0/286 (0%)
97.10G	87/190 (46%)	0/104 (0%)
109.8C	39/246 (16%)	4/183 (2%)
Overall (excluding control)	136/621 (22%)	4/573 (<1%)

[a] Numerators indicate the number of colonies hybridizing to the HIV-*tat* intron, while denominators indicate the total number of colonies from each plating. The percentage of HIV-*tat*-positive colonies is indicated in parentheses. The overall number and percentage of HIV-*tat* positive products from the three P1 experiments is shown in the bottom line.

Methods: Exon trapping was performed with three genomic P1 clones from human chromosome 16 (96.4B, 97.10G and 109.8C) and the corresponding no insert controls using either pSPL3 or pSPL3B as described in the legend for Fig. 1. The resulting PCR products were subcloned into pAMP10 (LifeTechnologies) using the UDG cloning method as described by the manufacturer (LifeTechnologies) followed by transformation into competent *E. coli* DH5α cells. Colony lifts were performed and the resulting filters were hybridized at 65°C in 0.25 M Na-phosphate pH 7.2/7% SDS, using the HIV-*tat* intron as a probe (described in the legend to Fig. 1). Filters were washed in 0.2 × SSC/0.1% SDS at 65°C for 30 min.

TABLE II

Plating efficiencies from ligation reactions of pSPL3-CAM

Cosmid	Subcloning efficiency		Ap^R/Cm^R ratio
	$Ap^{R a}$	$Cm^{R b}$	
pSPL3-CAM (no insert)	—	—	
LA2H2	6.5×10^7	1.8×10^6	36:1
LA325A11	9.5×10^7	2.2×10^6	43:1

[a] Transformation efficiencies based on colony-forming units per μg of pSPL3-CAM when plated on 50 μg Ap/ml.

[b] Transformation efficiencies based on colony-forming units per μg of pSPL3-CAM when plated on 25 μg Cm/ml.

Methods: The pSPL3-CAM vector was engineered by removing a 1.6-kb *Bam*HI fragment from pUC18CMR (ATCC No. 37718; Schweizer, 1990), which was then partially end-filled to leave a G overhang using PolIk. The resulting 1.6-kb Cm^R fragment was ligated with a 5023-bp *Bsp*HI fragment from pSPL3, which had been partially end-filled to leave a C overhang. *Bsp*HI digestion of pSPL3 cleanly excises the Ap^R cassette from the trapping vector. Cosmids LA325A11 and LA2H2 (containing genomic DNA from human chromosome 16) were digested with *Bam*HI + *Bgl*II and shotgun subcloned into *Bam*HI-cut and dephosphorylated pSPL3-CAM. The ligations were transformed into DH5α cells and plated onto LB agar plates containing 50 μg Ap/ml or 25 μg Cm/ml. Colonies were counted the following day. No colonies were seen on 'no insert' plates. Transformation efficiencies represent the average of two platings.

(data not shown). A simple solution to this problem would be to utilize restriction enzymes whose sites are present in the MCS of pSPL3 and cut within the Ap^R gene of cosmid vectors, or between the *ori* and the Ap^R gene in the cosmid. However, *Pst*I is the only enzyme in the pSPL3 MCS which cuts between the *ori* and the Ap^R gene in supercos or pWE cosmid vectors. To eliminate cosmid vector contamination and maximize the selection of usable sites in the pSPL3 MCS, the Ap^R gene of pSPL3 was replaced with the Cm^R gene as described in the legend to Table II. To examine the efficiency of this pSPL3 derivative, two cosmids were shotgun subcloned into pSPL3-CAM and the ligation reactions were plated on LB + Ap and LB + Cm plates. The results from these experiments are summarized in Table II. The cloning efficiencies of the cosmid-derived plasmids, based on the transformation efficiency on Ap plates, was 36–43-fold higher than that of pSPL3-CAM on Cm containing plates, confirming that the majority of ligation products are indeed cosmid derived. Colony hybridizations were performed on the resulting colonies using cosmid- or pSPL3-specific oligo probes. On filter lifts from the LB + Cm plates, every colony was seen to hybridize with oligo probes specific for the trapping vector, whereas less than 10% (3/39 and 5/47 for LA2H2 and LA325A11, respectively) of the colonies was seen to hybridize with the cosmid specific oligo (data not shown). Conversely, on filter lifts from the LB + Ap plates, more than 90% of the colonies from both reactions hybridized to a cosmid specific probe (data not shown). Based on this analysis, if pSPL3 had been used instead of pSPL3-CAM, only 2–3% of the resulting colonies would contain a trapping-vector-derived subclone. These data clearly demonstrate that replacement of Ap^R gene with the Cm^R gene produces a dramatic improvement in subcloning efficiency into the splicing vector.

(d) Conclusions and discussion

(*1*) In our initial studies using exon trapping, the majority of the trapped products were found to contain cryptic elements from the HIV-*tat* intron of the pSPL3 splicing vector. In this report, we have demonstrated that removal of a cryptic exon within the HIV-*tat* intron of pSPL3 significantly improves the efficiency of exon trapping. For all genomic P1 clones examined, the number of cryptic products was dramatically reduced when the modified trapping vector was compared to the original pSPL3 vector. In studies which are currently ongoing, we have utilized the pSPL3B vector to scan approximately 750 kb of genomic DNA for exons. In less than 1% of the cases examined did the trapped products hybridize to pSPL3B and not to the parental genomic P1 clone (data not shown).

(*2*) We have also demonstrated that replacement of the

Ap^R gene in pSPL3 with the Cm^R gene increases the initial subcloning efficiency of genomic cosmid clones by 36–43 fold. The presence of the Cm^R gene eliminates recircularized cosmid vector from the transformation products and insures that all the DNA used to transform the COS-7 cells is pSPL3 derived. As the method was originally described with an Ap^R splicing vector, as little as 2–3% of the DNA electroporated into COS cells would be pSPL3 recombinants. This may be sufficient for obtaining trapped exons; however, an increased likelihood of obtaining trapped exons can be achieved by increasing the initial subcloning efficiency by 36–43-fold with the pSPL3-CAM vector. We have recently combined both improvements to the pSPL3 vector by adding the Cm^R gene to the pSPL3B vector to produce pSPL3B-CAM.

(3) Church et al. (1994) found that approx. 4% of their trapped products contained HIV-*tat* sequences from the pSPL3 vector. However, using the pSPL3 exon-trapping vector as originally described, we found that on average, 20% of all trapped material was HIV-*tat* derived. In experiments using pools of plasmid minipreparations in which all the pSPL3 clones contained subcloned genomic DNAs, HIV-*tat* products represented up to 50% of the trapped exons examined (data not shown), suggesting that the HIV-*tat* products are independent of the initial subcloning efficiency and are not due to non-recombinant pSPL3 molecules being introduced into the COS cells. The level of HIV-*tat*-derived products was found to differ with each P1 or cosmid clone examined and may represent relative splicing efficiencies between the cryptic exon in pSPL3 and exons in the genomic subclones. The difference in abundance of the cryptic vector-derived exon reported by Church et al. (1994) may also be a consequence of clone-dependent variations in the relative splicing efficiency of pSPL3 subcloned DNA.

ACKNOWLEDGEMENTS

We would like to thank Timothy Houseal and Glenn Miller for critical review of the manuscript, David Munroe (MIT, Cambridge, MA, USA) for helpful suggestions during the course of this work and William Dackowski and Norman Doggett (Los Alamos National Laboratory, Los Alamos, NM, USA) for genomic P1 and cosmid clones, respectively. This work was supported by grant RO1 DK44853 to K.W.K.

REFERENCES

Aksentijevich, I., Pras, E., Gruberg, L., Shen, Y., Holman, K., Helling, S., Prosen, L., Sutherland, G.R., Richards, R.I., Ramsburg, M., Dean, M., Pras, M., Amos, C.I. and Kastner, D.L.: Refined mapping of the gene causing familial Mediterranean fever, by linkage and homozygosity studies. Am. J. Hum. Genet. 53 (1993) 451–461.

Auch, D. and Reth, M.: Exon trap cloning: using PCR to rapidly detect and clone exons from genomic DNA fragments. Nucleic Acids Res. 18 (1990) 6743–6744.

Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A. and Housman, D.E.: Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc. Natl. Acad. Sci. USA 88 (1991) 4005–4009.

Church, D.M., Banks, L.T., Rogers, A.C., Graw, S.L., Housman, D.E., Gusella, J.F. and Buckler, A.J.: Identification of human chromosome 9 specific genes using exon amplification. Human Mol. Genet. 2 (1993) 1915–1920.

Church, D.M., Stotler, C.J., Rutter, J.L., Murrell, J.R., Trofatter, J.A. and Buckler, A.J.: Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. Nature Genet. 6 (1994) 98–105.

Duyk, G.M., Kim, S.W., Myers, R.M. and Cox, D.R.: Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. Proc. Natl. Acad. Sci. USA 87 (1990) 8995–8999.

The European Chromosome 16 Tuberous Sclerosis Consortium: Identification and characterization of the tuberous sclerosis gene on chromosome 16. Cell 75 (1993) 1305–1315.

The European Polycystic Kidney Disease Consortium: The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. Cell 77 (1994) 881–894.

Feinberg, A.P. and Vogelstein, B.: A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132 (1983) 6–13.

Germino, G.G., Weinstat-Saslow, D., Himmelbauer, H., Gillespie, G.A.J., Somlo, S., Wirth, B., Barton, N., Harris, K.L., Frischauf, A.-M. and Reeders, S.T.: The gene for autosomal dominant polycystic kidney disease lies in a 750-kb CpG-rich region. Genomics 13 (1992) 144–151.

Mount, SM.: A catalogue of splice-junction sequences. Nucleic Acids Res. 10 (1982) 459–472.

Reeders, S.T., Breuning, M.H., Davies, K.E., Nicholls, R.D., Jarman, A.P., Higgs, D.R., Pearson, P.L. and Weatherall, D.J.: A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. Nature 317 (1985) 542–544.

Schweizer, H.P.: The pUC18M plasmids: a chloramphenicol resistance gene cassette for site-directed insertion and deletion mutagenesis in *Escherichia coli*. Biotechniques 8 (1990) 612–616.

Somlo, S., Wirth, B., Germino, G.G., Weinstat-Saslow, D., Gillespie, G.A.J., Himmelbauer, H., Steevens, L., Coucke, P., Willems, P., Bachner, L., Coto, E., Lopez-Larrea, C., Peral, B., San Millán, J.L., Saris, J.J., Breuning, M.H., Frischauf, A.-M. and Reeders, S.T.: Fine genetic localization of the gene for autosomal dominant polycystic kidney disease (PKD1) with respect to physically mapped markers. Genomics 13 (1992) 152–158.